Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 7 April 2003
Accepted 22 September 2003

The estimation of three-phase invariants when anomalous scatterers are present: the limits

Carmelo Giacovazzo,[a,b]* Massimo Ladisa[b] and Dritan Siliqi[a,c]

[a]Dipartimento Geomineralogico, Università di Bari, Campus Universitario, Via Orabona 4, 70125 Bari, Italy, [b]Istituto di Cristallografia, CNR, c/o Dipartimento Geomineralogico, Università di Bari, Campus Universitario, Via Orabona 4, 70125 Bari, Italy, and [c]Laboratory of X-ray Diffraction, Department of Inorganic Chemistry, Faculty of Natural Sciences, Tirana, Albania. Correspondence e-mail: carmelo.giacovazzo@ic.cnr.it

The probabilistic formula provided by Hauptman and Giacovazzo for estimating three-phase invariants when anomalous scatterers are present is revisited. Its main defects are: (*a*) it is absolutely resistant to any attempt at interpreting it in terms of parameters accessible *via* the experiment; (*b*) its calculation is time consuming and requires computing resources. A distribution based on interpretable estimates of the parameters is proposed. The role of the old and the new expressions in the single-wavelength anomalous diffraction (SAD) techniques is discussed, and compared with the role of analogous formulas estimating triplet invariants from isomorphous diffraction data.

1. Symbols and notation

$f = f' + if''$ atomic scattering factor: f' is its real part and may include an anomalous real effect, f'' is its imaginary part. The thermal factor is included.
$E_p = R_p \exp(i\varphi_p)$ = normalized structure factor of the native protein.
$E_d = R_d \exp(i\varphi_d)$ = normalized structure factor of the derivative.
$\Phi_p = \varphi_{p\mathbf{h}} + \varphi_{p\mathbf{k}} - \varphi_{p\mathbf{k}+\mathbf{h}}$.
N, a, na: number of non-hydrogen atoms in the protein unit cell, number of anomalous scatterers, number of non-anomalous scatterers respectively.
$E_a = R_a \exp(i\varphi_a)$ = normalized structure factor of the anomalous scatterer substructure.
I_0 modified Bessel function of order zero.
$\Sigma_H = \sum_j f_j^2$, where the summation is extended to the heavy atoms.
$[\sigma_i]_N, [\sigma_i]_H, [\sigma_i]_a = \sum_j Z_j^i$, where the summation is extended to all the protein atoms, to the heavy atoms only, to the anomalous scatterers, respectively (Z_j^i is the atomic number of the jth atom).
$\mathbf{h_1} = \mathbf{h}, \mathbf{h}_2 = \mathbf{k}, \mathbf{h}_3 = \overline{\mathbf{h}+\mathbf{k}}$.
$\Sigma_i = \sum_{j=1}^{N} [f_j'^2(\mathbf{h}_i) + f_j''^2]$.
$\Sigma_{ai} = \sum_{j=1}^{a} [f_j'^2(\mathbf{h}_i) + f_j''^2]$.
$\Sigma'_{ai} = \sum_{j=1}^{a} f_j'^2(\mathbf{h}_i)$.
$\Sigma''_{ai} = \sum_{j=1}^{a} [f_j'(\mathbf{h}_i) f_j'']$.
$\Sigma''_a = \sum_{j=1}^{a} f_j''^2$.
$\gamma_N = \sum_{j=1}^{N} f_j'(\mathbf{h}_1) f_j'(\mathbf{h}_2) f_j'(\mathbf{h}_3)/(\Sigma_1\Sigma_2\Sigma_3)^{1/2}$.
$\gamma_a = \sum_{j=1}^{a} f_j'(\mathbf{h}_1) f_j'(\mathbf{h}_2) f_j'(\mathbf{h}_3)/(\Sigma_1\Sigma_2\Sigma_3)^{1/2}$.
$\varepsilon = \left(\sum_1 \sum_2 \sum_3\right) \Big/ \left(\sum_{1na} \sum_{2na} \sum_{3na}\right)$.

2. Introduction

The probabilistic theory of the three-phase invariants for isomorphous pairs was initiated by Hauptman (1982*a*). He studied the distribution

$$P(E_{p\mathbf{h}}, E_{p\mathbf{k}}, E_{p\mathbf{h}+\mathbf{k}}, E_{d\mathbf{h}}, E_{d\mathbf{k}}, E_{d\mathbf{h}+\mathbf{k}}),$$

from which the conditional probability

$$P(\Phi_p | R_{p\mathbf{h}}, R_{p\mathbf{k}}, R_{p\mathbf{h}+\mathbf{k}}, R_{d\mathbf{h}}, R_{d\mathbf{k}}, R_{d\mathbf{h}+\mathbf{k}}) \approx [2\pi I_0(G)]^{-1} \exp(G\cos\Phi_p) \quad (1)$$

was derived. Equation (1) is a von Mises distribution: Φ_p is expected close to 0 or π according to whether G is positive or negative.

The algebraic expression of G is rather complicated (here not reported for brevity), and does not allow any easy interpretation in terms of parameters accessible *via* the diffraction experiment. The matter was revisited by Giacovazzo *et al.* (1988) who obtained for G the following simple expression:

$$G = 2[\sigma^3/\sigma_2^{3/2}]_p R_{p\mathbf{h}} R_{p\mathbf{k}} R_{p\mathbf{h}+\mathbf{k}} + 2[\sigma^3/\sigma_2^{3/2}]_H \Delta_{\mathbf{h}} \Delta_{\mathbf{k}} \Delta_{\mathbf{h}+\mathbf{k}}, \quad (2)$$

where

$$\Delta = (F_d - F_p)/\Sigma_H^{1/2}.$$

The Δs are isomorphous differences normalized with respect to the heavy-atom structure.

Expression (1) has been extensively tested by Furey *et al.* (1990), equation (2) has been the basis of a series of papers that describe how the protein phases may be estimated without any previous knowledge of the heavy-atom substructure (see Giacovazzo *et al.*, 1996, and literature quoted therein).

Estimates of the three-phase invariants when anomalous scatterers are present have been provided by Hauptman (1982*b*) and independently by Giacovazzo (1983). In spite of the quite different notation, the conclusive formulae provided by the two authors coincide. Their main result was the derivation of the joint probability distribution function

$$P(E_{\mathbf{h}}, E_{\mathbf{k}}, E_{\mathbf{h+k}}, E_{-\mathbf{h}}, E_{-\mathbf{k}}, E_{-\mathbf{h}-\mathbf{k}}) \equiv P(R_1, R_2, R_3, G_1, G_2, G_3, \varphi_1, \varphi_2, \varphi_3, \psi_1, \psi_2, \psi_3), \quad (3)$$

where R_i and G_i are normalized structure-factor moduli (*e.g.* $R = |F|/\Sigma^{1/2}$):

$$\begin{aligned} R_1 &= |E_{\mathbf{h}}|, & R_2 &= |E_{\mathbf{k}}|, & R_3 &= |E_{\mathbf{h+k}}|, \\ G_1 &= |E_{-\mathbf{h}}|, & G_2 &= |E_{-\mathbf{k}}|, & G_3 &= |E_{-\mathbf{h}-\mathbf{k}}|, \\ \varphi_1 &= \varphi_{\mathbf{h}}, & \varphi_2 &= \varphi_{\mathbf{k}}, & \varphi_3 &= \varphi_{\mathbf{h+k}}, \\ \psi_1 &= \varphi_{-\mathbf{h}}, & \psi_2 &= \varphi_{-\mathbf{k}}, & \psi_3 &= \varphi_{-\mathbf{h}-\mathbf{k}}. \end{aligned}$$

From equation (3), eight conditional distributions were derived:

$$P(\Phi_i|R_j, G_j, j = 1, 2, 3) \approx [2\pi I_0(\Omega_i)]^{-1} \exp[\Omega_i \cos(\Phi_i - \omega_i)], \quad (4)$$

for $i = 1, \ldots, 8$, where

$$\begin{aligned} \Phi_1 &= \psi_1 + \psi_2 + \varphi_3, & \Phi_2 &= \varphi_1 - \psi_2 - \varphi_3, \\ \Phi_3 &= \psi_1 - \varphi_2 + \varphi_3, & \Phi_4 &= \varphi_1 + \varphi_2 - \varphi_3, \\ \Phi_5 &= \psi_1 + \psi_2 - \psi_3, & \Phi_6 &= \varphi_1 - \psi_2 + \psi_3, \\ \Phi_7 &= \psi_1 - \varphi_2 - \psi_3, & \Phi_8 &= \varphi_1 + \varphi_2 + \psi_3. \end{aligned}$$

Equations (4) are von Mises distributions: they are unimodal, centred on ω_i, and Ω_i is the concentration parameter. As for the parameter G in (1), the algebraic expressions of the Ω_i's are absolutely impermeable to a straightforward interpretation in terms of parameters directly connected to the diffraction experiment.

In this paper, we intend to: (*a*) provide a probabilistic distribution for the triplet estimation immediately interpretable in terms of experimental parameters; (*b*) compare the efficiency of the new against the original expression; (*c*) describe the role of (4) in the direct procedures for phasing the proteins. In particular, we will show that (4) is expected to have a minor practical role compared with (1).

3. The simplified estimation of the parameters

3.1. Theoretical aspects

The estimates provided by (4), for $i = 1, \ldots, 8$, are strictly correlated with each other: thus only one of them is necessary for practical applications. We will focus our attention on the conditional distribution of Φ_4; we will simplify the notation by denoting Φ_4, Ω_4, ω_4, ϑ_4, γ_4 by Φ, Ω, ω, ϑ, γ, respectively. Accordingly, we rewrite the corresponding distribution as

$$P(\Phi|R_j, G_j, j = 1, 2, 3) \approx [2\pi I_0(\Omega)]^{-1} \exp[\Omega \cos(\Phi - \omega)], \quad (5)$$

where

Table 1
Numerical comparison between the parameters γ_s, ϑ_s, Ω_s, ω_s in the distribution (9) and γ, ϑ, Ω, ω in the distribution (4), for some selected triplets of TTG.

	(γ, γ_s)	(ϑ, ϑ_s)	(Ω, Ω_s)	(ω, ω_s)
Triplet 1	(−1.96, −2.28)	(−1.76, −1.98)	(2.63, 3.02)	(+48°, +49°)
Triplet 2	(−3.44, −3.53)	(−1.21, −1.21)	(3.65, 3.73)	(+71°, +71°)
Triplet 3	(+1.36, +1.33)	(−0.06, −0.07)	(1.36, 1.33)	(−87°, −87°)
Triplet 4	(+1.66, +1.61)	(−0.30, −0.30)	(1.69, 1.64)	(−78°, −79°)
Triplet 5	(+1.21, +1.11)	(−0.46, −0.47)	(1.29, 1.15)	(−69°, −67°)

$$\Omega = [\vartheta^2 + \gamma^2]^{1/2}, \quad (6)$$

$$\begin{aligned} \vartheta = {} & R_1R_2R_3A_4 \cos\xi_4 + G_1R_2R_3A_3D_1 \cos(\xi_3 - \xi_{01}) \\ & + R_1G_2R_3A_2D_2 \cos(\xi_2 + \xi_{02}) + R_1R_2G_3A_1D_3 \cos(\xi_1 - \xi_{03}) \\ & + R_1G_2G_3A_3D_2D_3 \cos(\xi_3 + \xi_{02} - \xi_{03}) \\ & + G_1R_2G_3A_2D_1D_3 \cos(\xi_2 - \xi_{01} + \xi_{03}) \\ & + G_1G_2R_3A_1D_1D_2 \cos(\xi_1 - \xi_{01} - \xi_{02}) \\ & + G_1G_2G_3A_4D_1D_2D_3 \cos(\xi_4 - \xi_{01} - \xi_{02} + \xi_{03}), \end{aligned} \quad (7)$$

$$\begin{aligned} \gamma = {} & R_1R_2R_3A_4 \sin\xi_4 - G_1R_2R_3A_3D_1 \sin(\xi_3 - \xi_{01}) \\ & + R_1G_2R_3A_2D_2 \sin(\xi_2 + \xi_{02}) + R_1R_2G_3A_1D_3 \sin(\xi_1 - \xi_{03}) \\ & + R_1G_2G_3A_3D_2D_3 \sin(\xi_3 + \xi_{02} - \xi_{03}) \\ & - G_1R_2G_3A_2D_1D_3 \sin(\xi_2 - \xi_{01} + \xi_{03}) \\ & - G_1G_2R_3A_1D_1D_2 \sin(\xi_1 - \xi_{01} - \xi_{02}) \\ & - G_1G_2G_3A_4D_1D_2D_3 \sin(\xi_4 - \xi_{01} - \xi_{02} + \xi_{03}), \end{aligned} \quad (8)$$

$$\omega = \tan^{-1}(\gamma/\omega).$$

For user usefulness, we collect in Appendix *A* the algebraic expressions of the various parameters that define the values of ϑ and γ and observe that:

(*a*) the terms A_{0i} and ξ_{0i}, for $i = 1, 2, 3$, can be expressed in terms of the coefficients c_{ji}, for $j = 1, 2$ and $i = 1, 2, 3$. In their turn, the c_{ji}'s are functions of the scattering factors.

(*b*) The coefficients A_i and ξ_i, for $i = 1, \ldots, 4$, can be expressed in terms of the parameters T_i, B_i, $i = 1, \ldots, 4$. In their turn, the T_i and B_i variables are functions of the quantities S_i, for $i = 1, \ldots, 8$, which depend on the values assumed by the variables Z_i, for $i = 1, \ldots, 8$. The latter depend on the variables t_i, b_i, $i = 1, \ldots, 4$, which are themselves complicated expressions depending on the chemical composition and on the used wavelength.

It is therefore not possible to deduce in a simple way, from the relationships (7) and (8), the expected value ω of the triplet invariant Φ and the reliability of the estimate.

The algebraic analysis of the various parameters allowed us to introduce a set of approximations that, for user usefulness, are collected in Appendix *B*. Such approximations enabled us to provide interpretable estimations of the parameters. Our final expression for the conditional probability distribution of the triplet phase invariants is the following:

$$P(\Phi|R_j, G_j, j = 1, 2, 3) \approx [2\pi I_0(\Omega_s)]^{-1} \exp[\Omega_s \cos(\Phi - \omega_s)], \quad (9)$$

where

$$\omega_s = \tan^{-1}(\gamma_s/\vartheta_s), \tag{10}$$

$$\gamma_s = -2[\sigma_3/\sigma_2^{3/2}]_a \Delta_1 \Delta_2 \Delta_3 m, \tag{11}$$

$$\vartheta_s = 2[\sigma_3/\sigma_2^{3/2}]_N \left\{ R_1R_2R_3 + 2R_3d_1d_2\left(1+\frac{1}{\tan\delta_1\tan\delta_2}\right) + 2R_2d_1d_3\left(1+\frac{1}{\tan\delta_1\tan\delta_3}\right) + 2R_1d_2d_3\left(1+\frac{1}{\tan\delta_2\tan\delta_3}\right) + d_1d_2d_3\left(1-\frac{1}{\tan\delta_1\tan\delta_2}+\frac{1}{\tan\delta_1\tan\delta_3}+\frac{1}{\tan\delta_2\tan\delta_3}\right)\right\}, \tag{12}$$

$$\Omega_s = (\gamma_s^2 + \vartheta_s^2)^{1/2}, \tag{13}$$

ω_s, γ_s, ϑ_s, Ω_s are simpler estimates of the parameters ω, γ, ϑ, Ω, respectively. ω_s is the estimated triplet invariant phase and Ω_s is its reliability. We have used the following notation:

$$\Delta_1 = \Delta_{\rm ano1}/\Sigma_{a1}^{1/2}, \quad \Delta_2 = \Delta_{\rm ano2}/\Sigma_{a2}^{1/2}, \quad \Delta_3 = \Delta_{\rm ano3}/\Sigma_{a3}^{1/2},$$

$$\Delta_{\rm ano1} = |F(\mathbf{h}_1)| - |F(-\mathbf{h}_1)|, \quad \Delta_{\rm ano2} = |F(\mathbf{h}_2)| - |F(-\mathbf{h}_2)|,$$

$$\Delta_{\rm ano3} = |F(\mathbf{h}_3)| - |F(-\mathbf{h}_3)|,$$

$$m = \left\{\prod_{i=1}^{3}[(\Sigma'_{ai}\Sigma_{ai})^{1/2}/\Sigma_{ai}^{m}]\right\}/8, \quad d_i = R_iG_i.$$

3.2. Analysis of the formula

Let us first analyse the algebraic expressions of the parameter γ_s. If we have only one type of anomalous scatterer, then

$$\begin{aligned} m &= \tfrac{1}{8}\prod_{i=1}^{3}\{[af'^2(h_i)a|f(h_i)|^2]^{1/2}/[af'(h_i)f'']\} \\ &= \tfrac{1}{8}\prod_{i=1}^{3}[|f(h_i)|/f''] \\ &= \prod_{i=1}^{3}1/(2\sin\delta_i) \end{aligned}$$

and

$$\gamma_s = -2[\sigma_3/\sigma_2^{3/2}]_a \Delta'_1\Delta'_2\Delta'_3, \tag{14}$$

where $\Delta'_i = \Delta_i/(2\sin\delta_i)$. The factor $(\sin\delta_i)^{-1}$ takes into account the larger signal provided by the anomalous scattering at higher $\sin\theta/\lambda$. The factor m changes also with the wavelength: it decreases when $\Delta f'$ becomes a larger negative number and/or when f'' increases.

Let us now analyse the term ϑ: it is a term of order $N^{-1/2}$ (while γ_s is of order $a^{-1/2}$). If the differences d_i are quite negligible with respect to R_i and to G_i, we can approximate ϑ_s by the Cochran contribution:

$$\vartheta_s \approx 2[\sigma_3/\sigma_2^{3/2}]_N R_1R_2R_3.$$

Expressions (10)–(13) clarify the role of the parameters accessible *via* the diffraction experiment. They were obtained under the hypothesis that only one type of anomalous scatterer is present. However, the relative order of magnitude of the B_i and T_i coefficients is not changed if more types of anomalous scatterers are in the unit cell. It may therefore be expected that (10)–(13) also hold under less strict conditions.

It is worthwhile interpreting in the Argand plane the coefficients G in (2) and Ω_s in (9). In the case of the isomorphous derivative, we define $\mathbf{G}$ as the sum of two real components (see Fig. 1*a*), the first always positive (say $v_1 = 2[\sigma_3/\sigma_2^{3/2}]_p R_1R_2R_3$) and the second positive or negative, and predominant (say $v_2 = [\sigma_3/\sigma_2^{3/2}]_H \Delta_{\mathbf{h}}\Delta_{\mathbf{k}}\Delta_{\mathbf{h}+\mathbf{k}}$). According to the relationships (10)–(13), $\mathbf{\Omega}_s$ may be defined as a vector (see Fig. 1*b*) having the real component ϑ_s defined by (12) and the imaginary component coinciding with γ_s as defined by (11). Ω_s is the modulus of $\mathbf{\Omega}_s$.

The following observations can be made:

(*a*) The real component is often negligible with respect to the imaginary component. Therefore the most reliable triplet phases are expected to have values close to $\pm\pi/2$.

(*b*) Negative (positive) values of γ should correspond to triplet phases close to $-\pi/2$ ($+\pi/2$).

(*c*) If the d_i's are negligible with respect to the R_i's and to the G_i's then $\cos\Phi$ is always expected to be positive.

(*d*) Our formulas (9)–(13) for SAD triplet invariants and formula (2) for isomorphous data triplets present a quite interesting similarity. The main difference between the two formulas may be expressed as follows: large values of $|\Delta_{\mathbf{h}}\Delta_{\mathbf{k}}\Delta_{\mathbf{h}+\mathbf{k}}|$ in (2) characterize triplet phases close to zero or π, large values of $|\Delta'_{\mathbf{h}}\Delta'_{\mathbf{k}}\Delta'_{\mathbf{h}+\mathbf{k}}|$ characterize triplet phases close to $\pm\pi/2$.

(*e*) The distribution (9) agrees well with the Karle (1984) first rule: 'if the sign of the product of the largest-magnitude differences ($\Delta_{\rm ano1}$, $\Delta_{\rm ano2}$, $\Delta_{\rm ano3}$) is the same as the sign of f'', the value of the average triplet invariant Φ is close to $-\pi/2$ and, when the signs are opposite, the value is close to $\pi/2$'. Formula (9) encompasses the Karle rule: indeed, it specifies the parameters defining the reliability of the Φ estimate and the sign of $\cos\Phi$.

3.3. Numerical analysis

In order to compare the efficiency of the original distribution (5) with the effectiveness of the distribution (9), we have selected two proteins:

(i) TTG (Walsh *et al.*, 1999), space group $C222_1$, a = 63.47, b = 65.96, c = 75.03 Å, 145 residues, 3 Se in the asymmetric unit, experimental data up to 2.28 Å resolution. Experimental

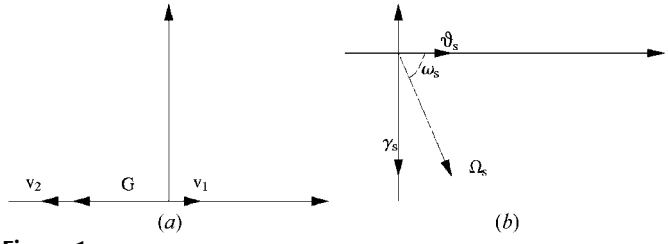

Figure 1
The concentration parameters G [see equation (2)] and Ω_s [see equation (13)] are interpreted in the Argand plane. (*a*) v_1 and v_2 are the two real components of G, where $v_1 = 2[\sigma_3/\sigma_2^{3/2}]_N R_1R_2R_3$ and $v_2 = [\sigma_3/\sigma_2^{3/2}]_H \Delta_{\mathbf{h}}\Delta_{\mathbf{k}}\Delta_{\mathbf{h}+\mathbf{k}}$. (*b*) ϑ_s, as given by equation (12) is the real component of Ω_s and $\gamma_s = [\sigma_3/\sigma_2^{3/2}]_a \Delta'_1\Delta'_2\Delta'_3$ is the imaginary component.

Table 2
Calculated data.

TTG and CauFd triplet invariants are found among the 920 and 860 reflections with the largest $|\Delta'|$ values, respectively; Ntr is the number of triplets with Ω or Ω_s larger than a given threshold Ω_{tr}, $\langle|\Delta\omega^0|\rangle$ and $\langle|\Delta\omega_s^0|\rangle$ are the corresponding phase errors [the first obtained *via* (5) and the second *via* (9)].

Structure	Ω_{tr}	Ntr	$\langle\|\Delta\omega^0\|\rangle$	Ntrs	$\langle\|\Delta\omega_s^0\|\rangle$
TTG	0.2	138183	66	128533	65
	0.4	49136	60	43836	59
	0.8	7051	51	6528	50
	1.2	1479	45	1413	44
	2.0	135	37	129	37
CauFd	0.2	18682	67	15187	67
	0.4	1150	57	1089	59
	0.8	44	51	39	52
	1.2	5	27	5	24

Table 3
Observed data.

TTG and CauFd triplet invariants are found among the 920 and 860 reflections with the largest experimental $|\Delta'|$ values; Ntr is the number of triplets with Ω or Ω_s larger than a given threshold Ω_{tr}, $\langle|\Delta\omega^0|\rangle$ and $\langle|\Delta\omega_s^0|\rangle$ are the corresponding phase errors [the first obtained *via* (5) and the second *via* (9)].

Structure	Ω_{tr}	Ntr	$\langle\|\Delta\omega^0\|\rangle$	Ntrs	$\langle\|\Delta\omega_s^0\|\rangle$
TTG	2.0	134335	83	133723	83
	4.4	68469	81	67451	81
	6.5	29737	79	29565	79
	15.0	1727	69	1731	70
CauFd	0.4	56690	83	55698	84
	1.2	6500	82	6250	84
	2.0	797	84	813	86
	4.4	27	107	26	108

data at four wavelengths are available for TTG, with the following expected parameters:

$\lambda_1 = 1.0781$ Å, $\Delta f_1' = -1.805$, $f_1'' = 0.646$,
$\lambda_2 = 0.9793$ Å, $\Delta f_2' = -8.852$, $f_2'' = 3.843$,
$\lambda_3 = 0.9791$ Å, $\Delta f_3' = -7.663$, $f_3'' = 3.841$,
$\lambda_4 = 0.9465$ Å, $\Delta f_4' = -2.618$, $f_4'' = 3.578$.

(ii) CauFd (Dauter *et al.*, 1997), space group $P4_32_12$, $a = 33.95$, $c = 74.82$ Å, 55 residues, 8 Fe in the asymmetric unit, experimental data up to 0.94 Å resolution. Experimental SAD data are available, at $\lambda = 0.88$ Å, with expected $\Delta f' = 0.26$ and $f'' = 1.25$.

We first checked the efficiency of the numerous approximations used in our mathematical approach. We show in Table 1 five triplets of TTG, with different resolutions and various parameter values, for which we compare the values of ω_s, γ_s, ϑ_s, Ω_s with the corresponding values of ω, γ, ϑ, Ω. The approximations prove to be sufficiently accurate for practical use (see also the tests presented in Tables 2 and 3).

We have then estimated triplet phases from calculated data by using $\Delta f_3'$ and f_3'' for TTG, and the above defined $\Delta f'$ and f'' values for CauFd. The results are shown in Table 2 (equivalent results are obtained for the other wavelengths): they indicate that (5) and (9) have equivalent efficiency. As a consequence, (9) may be considered a quite good approximation of (5): thus the first aim of this paper has been attained.

A surprise came out when we applied equations (5) and (9) to the corresponding experimental data (see Table 3). The measurement errors make (5) and (9) unuseful in practice: all triplets have reliability factors larger than 2.0 for TTG and larger than 0.4 for CauFd, but the corresponding phase errors are exceedingly high. We therefore decided to analyse the quality of the information carried by (5) and (9) to guess about their role in the practical phasing procedures.

4. The triplet invariant estimate when the anomalous scatterers are located

Suppose that at a certain stage of the phasing process the anomalous scatterer substructure is known. Then the structure-factor phases can be estimated, for example *via* distributions like that of Giacovazzo, Ladisa & Siliqi (2002):

$$P(\varphi^+|R, G, E_a^+, E_a^-) \approx [2\pi I_o(X)]^{-1} \exp[X\cos(\varphi^+ - \vartheta^+)], \quad (15)$$

where

$$E^+ = R\exp(i\varphi^+), \quad E^- = G\exp(i\varphi^-),$$
$$E_a^+ = R_a\exp(i\varphi_a^+), \quad E_a^- = G_a\exp(i\varphi_a^-)$$

are pseudo-normalized structure factors [normalized with respect to the non-anomalous scatterer substructure; *e.g.* $E^+ = F^+/(\sum_{na})^{1/2}$] ,

$$\tan\vartheta^+ = P/Q, \quad (16)$$
$$P = 2(q^+RR_a\sin\varphi_a^+ - q^-GG_a\sin\varphi_a^-) + 2\frac{(R-G)}{e}[R_a\sin\varphi_a^+ + G_a\sin\varphi_a^-],$$
$$Q = 2(q^+RR_a\cos\varphi_a^+ + q^-GG_a\cos\varphi_a^-) + 2\frac{(R-G)}{e}[R_a\cos\varphi_a^+ - G_a\cos\varphi_a^-],$$
$$X = (P^2+Q^2)^{1/2}, \quad (17)$$
$$e = (\langle|\mu^+|^2\rangle + \langle|\mu^-|^2\rangle)/\sum_{na}.$$

μ^+ and μ^- represent the cumulative errors arising from different sources (*i.e.* the structural model constituted by the located anomalous scatterers and errors in measurements).

Both P and Q have two contributors: the first is a Sim-like term (Sim, 1959, 1960), the second depends on the Δ_{ano} experimental measures. If the Sim contribution is neglected (*e.g.* for big structures and/or for small values of R_a and G_a), then

$$X = 4\Delta_{ano}|F_a''^+|/[|\mu^+|^2 + |\mu^-|^2],$$
$$\vartheta^+ = \varphi_a''^+ + \mathrm{sign}(\Delta_{ano})\pi/2, \quad (18)$$

where $\varphi_a''^+$ is the phase of

$$F_a''^+ = \sum_{j=1}^{a} f_j''\exp(2\pi i\mathbf{h}\cdot\mathbf{r}_j).$$

Let us now apply (15) to the calculated and experimental data of TTG and CauFd. We use the same reflections employed in

Tables 2 and 3. The results are summarized in Table 4. The overall phase error ($|Er|^\circ$) is small even for experimental data: the formula (15) is therefore useful for practical applications.

A question arises: can we identify the information lost by (9) and exploited by (15)? To this aim, we verify how the information on the anomalous-scatterer substructure modifies the triplet-invariant estimates. The necessary tool for obtaining the new estimates is the knowledge of the conditional distribution

$$P(\Phi|\{R_i, G_i, E^+_{ai}, E^-_{ai}, i = 1, 2, 3\}). \quad (19)$$

The distribution (19) may be derived *via* a mathematical approach similar to that used by Giacovazzo, Siliqi & De Caro (2002) to estimate triplet-phase invariants in the isomorphous-replacement case when the heavy-atom substructure is known. In particular, we will consider (19) as the combination of the three independent distributions (15), where $i = 1, 2, 3$ refers to **h**, **k**, **h** + **k**, respectively:

$$P(\Phi|\{R_i, G_i, E^+_{ai}, E^-_{ai}, i = 1, 2, 3\}) \approx \exp\{\Sigma_i X_i \cos(\varphi^+_i - \vartheta^+_i)\}. \quad (20)$$

We obtain

$$P(\Phi|\{R_i, G_i, E^+_{ai}, E^-_{ai}, i = 1, 2, 3\}) \approx [2\pi I_0(\tau)]^{-1} \exp[\tau \cos(\Phi - \Theta)], \quad (21)$$

where

$$\Theta = \vartheta^+_1 + \vartheta^+_2 - \vartheta^+_3$$

and τ is defined by

$$D_1(\tau) = D_1(X_1)D_1(X_2)D_1(X_3), \quad (22)$$

where $D_1(x) = I_1(x)/I_0(X)$.

Equation (21) is the required expression: it should provide improved estimates of the triplet invariants [with respect to those given by (5) or by (9)]. The main behaviour of (21) may be discovered by assuming that the anomalous scattering of the protein is small with respect to the normal scattering. Then, in accordance with equation (18),

$$X_i = 4\Delta_{\mathrm{ano}i}|F''^+_{ai}|/[\langle|\mu^+_i|^2\rangle + \langle|\mu^-_i|^2\rangle], \quad i = 1, 2, 3.$$

If the X_i's are relatively small (say <0.6), then $D_1(X_i) \approx X_i/2$,

$$\tau = 2\frac{|\Delta_{\mathrm{ano1}}|}{\langle|\mu^+_1|^2\rangle + \langle|\mu^-_1|^2\rangle}\frac{|\Delta_{\mathrm{ano2}}|}{\langle|\mu^+_2|^2\rangle + \langle|(\mu^-_2)^2|\rangle} \times \frac{|\Delta_{\mathrm{ano3}}|}{\langle|\mu^+_3|^2\rangle + \langle|(\mu^-_3)^2|\rangle}|F''_{a1}F''_{a2}F''_{a3}| \quad (23)$$

and

$$\Theta = \Phi''^+_a + \pi/2[\mathrm{sign}(\Delta_{\mathrm{ano1}}) + \mathrm{sign}(\Delta_{\mathrm{ano2}}) + \mathrm{sign}(\Delta_{\mathrm{ano3}})], \quad (24)$$

where

$$\Phi''^+_a = \varphi''^+_1 + \varphi''^+_2 - \varphi''^+_3.$$

If the number of anomalous scatterers is small, Φ''^+_a is expected to be close to zero (it is just this expectation that makes the triplet phases estimable in the absence of information on the anomalous-scatterer substructure). In this case, the phase estimates *via* (24) are nearly equivalent to those provided by (9): the only difference is that the estimates (24) are strengthened if the product $|F''_{a1}F''_{a2}F''_{a3}|$ is large. The reader can easily verify that, when $\Phi''^+_a \approx 0$, (24) and (9) provide the following estimates:

if $\Delta_{\mathrm{ano1}} > 0$, $\Delta_{\mathrm{ano2}} > 0$, $\Delta_{\mathrm{ano3}} > 0$ then $\omega_s = \Theta = -\pi/2$
if $\Delta_{\mathrm{ano1}} > 0$, $\Delta_{\mathrm{ano2}} > 0$, $\Delta_{\mathrm{ano3}} < 0$ then $\omega_s = \Theta = +\pi/2$
if $\Delta_{\mathrm{ano1}} < 0$, $\Delta_{\mathrm{ano2}} < 0$, $\Delta_{\mathrm{ano3}} > 0$ then $\omega_s = \Theta = -\pi/2$
if $\Delta_{\mathrm{ano1}} < 0$, $\Delta_{\mathrm{ano2}} < 0$, $\Delta_{\mathrm{ano3}} < 0$ then $\omega_s = \Theta = +\pi/2$.

It may be worthwhile stressing that, if Φ''^+_a is not close to zero, the distributions (24) and (9) will provide different estimates. Unfortunately, this is just the case when the triplets are found among the reflections with the largest $|\Delta'|$ values (the most reliable ones): indeed, a large $|\Delta'|$ difference does not imply large R or G values. To give a numerical example, the average values of $|\Phi''^+_a|$ for the triplets analysed in Table 1 are:

$$\langle|\Phi''^+_a|\rangle = 51^\circ \text{ for TTG}, \quad \langle|\Phi''^+_a|\rangle = 72^\circ \text{ for CauFd}.$$

To verify how the triplet estimates improve when the information on the anomalous substructure (and therefore on Φ''^+_a) is available, we have applied the distribution (21) to the calculated and observed data of TTG and CauFd: we used the same reflections employed in Tables 2 and 3.

For the calculated data, we obtained:

for TTG: 130306 triplets with $\tau > 0.2$ and average phase error $\langle|\Delta\omega^0|\rangle = 45^\circ$;

for CauFd: 107435 triplets with $\tau > 0.2$ and $\langle|\Delta\omega^0|\rangle = 54^\circ$.

For the observed data, the outcome was the following:

for TTG: 128060 triplets with $\tau > 0.4$ and $\langle|\Delta\omega^0|\rangle = 78^\circ$;

for CauFd: 38713 triplets with $\tau > 0.4$ and $\langle|\Delta\omega^0|\rangle = 59^\circ$ (the quality of the CauFd experimental data is superior to that of TTG).

The triplet estimates including the information on the substructure are markedly better than estimates without it, but worse than the estimates on the phases of the single reflections. Furthermore, the experimental errors deteriorate the quality of the information (the errors on the single Δ_{ano} sum in the triplet expression).

Table 4
Phase errors obtained *via* the probability distribution function (15), when applied to the calculated and to the experimental data of TTG and CauFd.

Structure	Calc. data $|Er|^\circ$	Exp. data $|Er|^\circ$
TTG	26	54
CauFd	33	38

5. Conclusions

The above results allow us to formulate the following considerations:

(*a*) Distribution (5) is largely less informative than distribution (15) (which requires and exploits information on the

anomalous substructure) even in the absence of experimental errors.

(*b*) The efficiency of (5) is strongly deteriorated by the unavoidable experimental errors [formula (5) does not take the errors into account and therefore the phase reliability is strongly overestimated]. On the contrary, (15) treats the errors as supplementary primitive variables and therefore is more robust.

(*c*) Attempts at finding directly the protein phases from (5), or equivalently from (9), are discouraged. This conclusion is in contrast with the procedure recently suggested by Giacovazzo, Ladisa & Siliqi (2002) for the isomorphous-replacement case (the triplet-invariant estimates are first used to evaluate a subset of protein phases, from which the heavy-atom substructure may be routinely identified).

(*d*) The additional reason for the inefficiency of (5) is the fact that it assumes that $\Phi_a''^{+}$ is close to zero. The practice of introducing Se atoms into a protein as selenomethionines makes frequent the cases in which $a > 20$ (cases occur for which $a > 100$). In all these cases, the $\Phi_a''^{+}$ values will be significantly different from zero even if $|E_{ai}^{+}|$, for $i = 1, 2, 3$, are large.

All the above considerations allow us to predict that equations like (5), or equivalently (9), are expected to have a marginal role in SAD procedures.

APPENDIX *A*

We collect in this Appendix the algebraic expressions of the parameters defining the values of the variables Ω and ω in equation (5).

$$c_{1i} = \sum_{j=1}^{N}[f_j'^2(\mathbf{h}_i) - f_j''^2]/\Sigma_i, \quad c_{2i} = 2\sum_{j=1}^{N}[f_j'(\mathbf{h}_i)f_j'']/\Sigma_i,$$
$$c_i = [1 - (c_{1i}^2 + c_{2i}^2)]^2, \quad i = 1, 2, 3;$$

$$\sin\xi_{0i} = c_{2i}/(c_{1i}^2 + c_{2i}^2)^{1/2}, \quad \cos\xi_{0i} = c_{1i}/(c_{1i}^2 + c_{2i}^2)^{1/2},$$
$$A_{0i} = 2[(c_{1i}^2 + c_{2i}^2)/c_i]^{1/2}, \quad i = 1, 2, 3;$$

$$t_1 = 1/(\Sigma_1\Sigma_2\Sigma_3)^{1/2}\sum_{j=1}^{N}\{f_j'(\mathbf{h}_1)f_j'(\mathbf{h}_2)f_j'(\mathbf{h}_3) - f_j''^2[f_j'(\mathbf{h}_1) + f_j'(\mathbf{h}_2) + f_j'(\mathbf{h}_3)]\};$$
$$t_2 = 1/(\Sigma_1\Sigma_2\Sigma_3)^{1/2}\sum_{j=1}^{N}\{f_j'(\mathbf{h}_1)f_j'(\mathbf{h}_2)f_j'(\mathbf{h}_3) + f_j''^2[-f_j'(\mathbf{h}_1) + f_j'(\mathbf{h}_2) + f_j'(\mathbf{h}_3)]\};$$
$$t_3 = 1/(\Sigma_1\Sigma_2\Sigma_3)^{1/2}\sum_{j=1}^{N}\{f_j'(\mathbf{h}_1)f_j'(\mathbf{h}_2)f_j'(\mathbf{h}_3) + f_j''^2[f_j'(\mathbf{h}_1) - f_j'(\mathbf{h}_2) + f_j'(\mathbf{h}_3)]\};$$
$$t_4 = 1/(\Sigma_1\Sigma_2\Sigma_3)^{1/2}\sum_{j=1}^{N}\{f_j'(\mathbf{h}_1)f_j'(\mathbf{h}_2)f_j'(\mathbf{h}_3) + f_j''^2[f_j'(\mathbf{h}_1) + f_j'(\mathbf{h}_2) - f_j'(\mathbf{h}_3)]\};$$

$$b_1 = 1/(\Sigma_1\Sigma_2\Sigma_3)^{1/2}\sum_{j=1}^{N}\{f_j''[f_j'(\mathbf{h}_1)f_j'(\mathbf{h}_2) + f_j'(\mathbf{h}_1)f_j'(\mathbf{h}_3) + f_j'(\mathbf{h}_2)f_j'(\mathbf{h}_3)] - f_j''^3\};$$
$$b_2 = 1/(\Sigma_1\Sigma_2\Sigma_3)^{1/2}\sum_{j=1}^{N}\{f_j''[f_j'(\mathbf{h}_1)f_j'(\mathbf{h}_2) + f_j'(\mathbf{h}_1)f_j'(\mathbf{h}_3) - f_j'(\mathbf{h}_2)f_j'(\mathbf{h}_3)] + f_j''^3\};$$
$$b_3 = 1/(\Sigma_1\Sigma_2\Sigma_3)^{1/2}\sum_{j=1}^{N}\{f_j''[f_j'(\mathbf{h}_1)f_j'(\mathbf{h}_2) - f_j'(\mathbf{h}_1)f_j'(\mathbf{h}_3) + f_j'(\mathbf{h}_2)f_j'(\mathbf{h}_3)] + f_j''^3\};$$
$$b_4 = 1/(\Sigma_1\Sigma_2\Sigma_3)^{1/2}\sum_{j=1}^{N}\{f_j''[-f_j'(\mathbf{h}_1)f_j'(\mathbf{h}_2) + f_j'(\mathbf{h}_1)f_j'(\mathbf{h}_3) + f_j'(\mathbf{h}_2)f_j'(\mathbf{h}_3)] + f_j''^3\};$$

$$Z_1 = (t_4 - c_{11}t_3 - c_{21}b_3)/(c_1)^{1/2};$$
$$Z_2 = (t_1 - c_{11}t_2 + c_{21}b_2)/(c_1)^{1/2};$$
$$Z_3 = (t_3 - c_{11}t_4 - c_{21}b_4)/(c_1)^{1/2};$$
$$Z_4 = (t_2 - c_{11}t_1 - c_{21}b_1)/(c_1)^{1/2};$$
$$Z_5 = (b_4 + c_{11}b_3 - c_{21}t_3)/(c_1)^{1/2};$$
$$Z_6 = (b_1 - c_{11}b_2 - c_{21}t_2)/(c_1)^{1/2};$$
$$Z_7 = (b_3 + c_{11}b_4 - c_{21}t_4)/(c_1)^{1/2};$$
$$Z_8 = (-b_2 + c_{11}b_1 - c_{21}t_1)/(c_1)^{1/2};$$

$$S_1 = (-Z_2 + c_{12}Z_3 - c_{22}Z_7)/(c_2)^{1/2};$$
$$S_2 = (-Z_1 + c_{12}Z_4 - c_{22}Z_8)/(c_2)^{1/2};$$
$$S_3 = (-Z_6 + c_{12}Z_7 + c_{22}Z_3)/(c_2)^{1/2};$$
$$S_4 = (-Z_5 + c_{12}Z_8 + c_{22}Z_4)/(c_2)^{1/2};$$
$$S_5 = (-Z_4 + c_{12}Z_1 + c_{22}Z_5)/(c_2)^{1/2};$$
$$S_6 = (-Z_3 + c_{12}Z_2 + c_{22}Z_6)/(c_2)^{1/2};$$
$$S_7 = (-Z_8 + c_{12}Z_5 - c_{22}Z_1)/(c_2)^{1/2};$$
$$S_8 = (-Z_7 + c_{12}Z_6 - c_{22}Z_2)/(c_2)^{1/2};$$

$$T_1 = 2(-S_1 + c_{13}S_2 - c_{23}S_4)/(c_3)^{1/2};$$
$$T_2 = 2(-S_5 + c_{13}S_6 + c_{23}S_8)/(c_3)^{1/2};$$
$$T_3 = 2(-S_6 + c_{13}S_5 - c_{23}S_7)/(c_3)^{1/2};$$
$$T_4 = 2(-S_2 + c_{13}S_1 + c_{23}S_3)/(c_3)^{1/2};$$
$$B_1 = 2(-S_3 + c_{13}S_4 + c_{23}S_2)/(c_3)^{1/2};$$
$$B_2 = 2(-S_7 + c_{13}S_8 - c_{23}S_6)/(c_3)^{1/2};$$
$$B_3 = 2(-S_8 + c_{13}S_7 + c_{23}S_5)/(c_3)^{1/2};$$
$$B_4 = 2(-S_4 + c_{13}S_3 - c_{23}S_1)/(c_3)^{1/2};$$

$$A_i = (T_i^2 + B_i^2)^{1/2}, \quad \sin\xi_i = B_i/(T_i^2 + B_i^2)^{1/2},$$
$$\cos\xi_i = T_i/(T_i^2 + B_i^2)^{1/2}, \quad i = 1, 2, 3, 4.$$

APPENDIX *B*

We collect in this Appendix the basic approximations aiming at providing simple estimates of the parameters defining the distribution (5).

We first simplify the parameters ϑ, γ in (5) as follows:

$$\vartheta \approx R_1R_2R_3T_4 + G_1R_2R_3T_3 + R_1G_2R_3T_2 + R_1R_2G_3T_1 + R_1G_2G_3T_3 + G_1R_2G_3T_2 + G_1G_2R_3T_1 + G_1G_2G_3T_4 \quad (25)$$

$$\gamma \approx R_1R_2R_3B_4 - G_1R_2R_3B_3 + R_1G_2R_3B_2 + R_1R_2G_3B_1 + R_1G_2G_3B_3 - G_1R_2G_3B_2 - G_1G_2R_3B_1 - G_1G_2G_3B_4 \quad (26)$$

$\omega \approx \tan^{-1}(\gamma/\vartheta)$.

To obtain (25) and (26):

(*a*) We have assumed $\xi_{0i} = 0$ for $i = 1, 2, 3$. In practice, these angles are very small, as one can directly assess from the definitions given in Appendix *A*. Indeed, $c_{2i} \ll c_{1i}$.

(*b*) We have assumed $D_i = 1$ for $i = 1, 2, 3$, where

$$D_1 = D(A_{01}R_1G_1), \quad D_2 = D(A_{02}R_2G_2), \quad D_3 = D(A_{03}R_3G_3)$$

and

$$D(x) = I_1(x)/I_0(x).$$

Accordingly, we focus our interest on the triplet invariants with large values of the products R_iG_i, for i = 1, 2, 3. They belong to the subset potentially most useful to derive structure information *via* electron-density maps.

(*c*) We have replaced the variables (A_i, ξ_i) of the original formula by the variables (T_i, B_i), in accordance with Appendix *A*.

To simplify further the relationships (25)–(26), we note that, according to the definitions given in Appendix *A*,

$$(1 - c_{1i})\Sigma_i = 2\Sigma_a'', \quad c_{2i}\Sigma_i = 2\Sigma_{ai}^m \quad (i = 1, 2, 3),$$

from which

$$(1 - c_{1i})/c_{2i} = \Sigma_a''/\Sigma_{ai}^m \quad (i = 1, 2, 3).$$

If we assume that only one type of anomalous scatterer is present, then

$$(1 - c_{1i})/c_{2i} = f''/f'(\mathbf{h}_i) = \tan\delta_i, \quad \text{for } i = 1, 2, 3,$$

where δ_i is, in the Argand plane, the angle of the scattering factor of anomalous scatterer [*i.e.* $f = |f|\exp(i\delta)$]. Note that δ varies with the resolution.

We can now rewrite the variables b_i and t_i, i = 1, ..., 4, defined in Appendix *A*, as follows:

$$t_1 \cong \gamma_N - \gamma_a\left(\frac{1-c_{12}}{c_{22}}\frac{1-c_{13}}{c_{23}} + \frac{1-c_{11}}{c_{21}}\frac{1-c_{13}}{c_{23}} + \frac{1-c_{12}}{c_{22}}\frac{1-c_{11}}{c_{21}}\right),$$

$$t_2 \cong \gamma_N + \gamma_a\left(-\frac{1-c_{12}}{c_{22}}\frac{1-c_{13}}{c_{23}} + \frac{1-c_{11}}{c_{21}}\frac{1-c_{13}}{c_{23}} + \frac{1-c_{12}}{c_{22}}\frac{1-c_{11}}{c_{21}}\right),$$

$$t_3 \cong \gamma_N + \gamma_a\left(\frac{1-c_{12}}{c_{22}}\frac{1-c_{13}}{c_{23}} - \frac{1-c_{11}}{c_{21}}\frac{1-c_{13}}{c_{23}} + \frac{1-c_{12}}{c_{22}}\frac{1-c_{11}}{c_{21}}\right),$$

$$t_4 \cong \gamma_N + \gamma_a\left(\frac{1-c_{12}}{c_{22}}\frac{1-c_{13}}{c_{23}} + \frac{1-c_{11}}{c_{21}}\frac{1-c_{13}}{c_{23}} - \frac{1-c_{12}}{c_{22}}\frac{1-c_{11}}{c_{21}}\right),$$

$$b_1 \cong \gamma_a\left(\frac{1-c_{13}}{c_{23}} + \frac{1-c_{12}}{c_{22}} + \frac{1-c_{11}}{c_{21}} - \frac{1-c_{13}}{c_{23}}\frac{1-c_{12}}{c_{22}}\frac{1-c_{11}}{c_{21}}\right),$$

$$b_2 \cong \gamma_a\left(\frac{1-c_{13}}{c_{23}} + \frac{1-c_{12}}{c_{22}} - \frac{1-c_{11}}{c_{21}} + \frac{1-c_{13}}{c_{23}}\frac{1-c_{12}}{c_{22}}\frac{1-c_{11}}{c_{21}}\right),$$

$$b_3 \cong \gamma_a\left(\frac{1-c_{13}}{c_{23}} - \frac{1-c_{12}}{c_{22}} + \frac{1-c_{11}}{c_{21}} + \frac{1-c_{13}}{c_{23}}\frac{1-c_{12}}{c_{22}}\frac{1-c_{11}}{c_{21}}\right),$$

$$b_4 \cong \gamma_a\left(-\frac{1-c_{13}}{c_{23}} + \frac{1-c_{12}}{c_{22}} + \frac{1-c_{11}}{c_{21}} + \frac{1-c_{13}}{c_{23}}\frac{1-c_{12}}{c_{22}}\frac{1-c_{11}}{c_{21}}\right).$$

We stress that the formulas written above exactly hold in the limit of a single type of anomalous scatterer.

If we rewrite the expressions (defined in Appendix *A*) of the variables Z_i and S_i, i = 1, ..., 8, and introduce them in the algebraic definitions of the variables B_i and T_i, which fix the values of ϑ and γ in (25) and (26), we obtain:

$$B_1 \cong -\frac{2}{c_{21}c_{22}c_{23}}\gamma_a + \frac{\varepsilon}{4}\left(\frac{c_{21}}{1-c_{11}} - \frac{c_{22}}{1-c_{12}} - \frac{c_{23}}{1-c_{13}} - \frac{c_{21}}{1-c_{11}}\frac{c_{22}}{1-c_{12}}\frac{c_{23}}{1-c_{13}}\right)\gamma_a - \frac{\varepsilon}{4}\left(\frac{c_{21}}{1-c_{11}} + \frac{c_{22}}{1-c_{12}} + \frac{c_{23}}{1-c_{13}} - \frac{c_{21}}{1-c_{11}}\frac{c_{22}}{1-c_{12}}\frac{c_{23}}{1-c_{13}}\right)\gamma_N,$$

$$B_2 \cong -\frac{2}{c_{21}c_{22}c_{23}}\gamma_a + \frac{\varepsilon}{4}\left(\frac{c_{21}}{1-c_{11}} - \frac{3c_{22}}{1-c_{12}} - \frac{3c_{23}}{1-c_{13}} - \frac{c_{21}}{1-c_{11}}\frac{c_{22}}{1-c_{12}}\frac{c_{23}}{1-c_{13}}\right)\gamma_a - \frac{\varepsilon}{4}\left(\frac{c_{21}}{1-c_{11}} - \frac{c_{22}}{1-c_{12}} - \frac{c_{23}}{1-c_{13}} - \frac{c_{21}}{1-c_{11}}\frac{c_{22}}{1-c_{12}}\frac{c_{23}}{1-c_{13}}\right)\gamma_N,$$

$$B_3 \cong \frac{2}{c_{21}c_{22}c_{23}}\gamma_a + \frac{\varepsilon}{4}\left(\frac{c_{21}}{1-c_{11}} + \frac{c_{22}}{1-c_{12}} + \frac{3c_{23}}{1-c_{13}} + \frac{c_{21}}{1-c_{11}}\frac{c_{22}}{1-c_{12}}\frac{c_{23}}{1-c_{13}}\right)\gamma_a - \frac{\varepsilon}{4}\left(\frac{c_{21}}{1-c_{11}} - \frac{c_{22}}{1-c_{12}} + \frac{c_{23}}{1-c_{13}} + \frac{c_{21}}{1-c_{11}}\frac{c_{22}}{1-c_{12}}\frac{c_{23}}{1-c_{13}}\right)\gamma_N,$$

$$B_4 \cong \frac{2}{c_{21}c_{22}c_{23}}\gamma_a + \frac{\varepsilon}{4}\left(\frac{c_{21}}{1-c_{11}} + \frac{3c_{22}}{1-c_{12}} + \frac{c_{23}}{1-c_{13}} + \frac{c_{21}}{1-c_{11}}\frac{c_{22}}{1-c_{12}}\frac{c_{23}}{1-c_{13}}\right)\gamma_a - \frac{\varepsilon}{4}\left(\frac{c_{21}}{1-c_{11}} + \frac{c_{22}}{1-c_{12}} - \frac{c_{23}}{1-c_{13}} + \frac{c_{21}}{1-c_{11}}\frac{c_{22}}{1-c_{12}}\frac{c_{23}}{1-c_{13}}\right)\gamma_N,$$

$$T_1 \cong \frac{\varepsilon}{4}(\gamma_N - \gamma_a)(1 + P_3), \quad T_2 \cong \frac{\varepsilon}{4}(\gamma_N - \gamma_a)(1 + P_2),$$

$$T_3 \cong \frac{\varepsilon}{4}(\gamma_N - \gamma_a)(1 + P_1), \quad T_4 \cong \frac{\varepsilon}{4}(\gamma_N - \gamma_a)(1 + P_4),$$

where

$$P_1 \cong \frac{c_{21}}{1-c_{11}}\frac{c_{22}}{1-c_{12}} - \frac{c_{21}}{1-c_{11}}\frac{c_{23}}{1-c_{13}} + \frac{c_{22}}{1-c_{12}}\frac{c_{23}}{1-c_{13}},$$

$$P_2 \cong \frac{c_{21}}{1-c_{11}}\frac{c_{22}}{1-c_{12}} + \frac{c_{21}}{1-c_{11}}\frac{c_{23}}{1-c_{13}} - \frac{c_{22}}{1-c_{12}}\frac{c_{23}}{1-c_{13}},$$

$$P_3 \cong -\frac{c_{21}}{1-c_{11}}\frac{c_{22}}{1-c_{12}} - \frac{c_{21}}{1-c_{11}}\frac{c_{23}}{1-c_{13}} - \frac{c_{22}}{1-c_{12}}\frac{c_{23}}{1-c_{13}},$$

$$P_4 \cong -\frac{c_{21}}{1-c_{11}}\frac{c_{22}}{1-c_{12}} + \frac{c_{21}}{1-c_{11}}\frac{c_{23}}{1-c_{13}} + \frac{c_{22}}{1-c_{12}}\frac{c_{23}}{1-c_{13}}.$$

Let us analyse the order of magnitude of the three terms contributing to the B_i's. We observe:

(*a*) c_{2i} is of order a/N;

(*b*) although both c_{2i} and $(1 - c_{1i})$ are small quantities, their ratio is finite and of order 1;

(*c*) ε is close to unity;

(*d*) γ_N is of order $N^{-1/2}$ and γ_a is of order $a/N^{3/2}$.

Accordingly, the first term in B_i is of order $N^{3/2}/a^2$, usually a large number. The second and third terms are of order $a/N^{3/2}$ and $N^{-1/2}$, respectively, and therefore are both negligible with respect to the first one.

Accordingly, the following relationship can be established:

$$B_1 \approx B_2 \approx -B_3 \approx B_4 \approx -2\gamma_a/(c_{21}c_{22}c_{23}). \quad (27)$$

Putting (27) into (26) and renormalizing R and G with respect to the anomalous-scatterers structure gives

$$\gamma \approx -2[\sigma_3/\sigma_2^{3/2}]_a \Delta_1 \Delta_2 \Delta_3 m. \quad (28)$$

Let us now analyse the terms T_i. We neglect γ_a with respect to γ_N and, assuming $\varepsilon \approx 1$, we have

$$\frac{\varepsilon}{4}(\gamma_N - \gamma_a) \cong \tfrac{1}{4}[\sigma_3/\sigma_2^{3/2}]_N.$$

Accordingly, ϑ becomes

$$\vartheta = 1/4[\sigma_3/\sigma_2^{3/2}]_N(u+v),$$

where

$$\begin{aligned} u &= R_1R_2R_3 + G_1R_2R_3 + R_1G_2R_3 + R_1R_2G_3 + R_1G_2G_3 \\ &\quad + G_1R_2G_3 + G_1G_2R_3 + G_1G_2G_3, \\ v &= (R_1R_2R_3 + G_1G_2G_3)P_4 + (G_1R_2R_3 + R_1G_2G_3)P_1 \\ &\quad + (R_1G_2R_3 + G_1R_2G_3)P_2 + (R_1R_2G_3 + G_1G_2R_3)P_3. \end{aligned} \quad (29)$$

If the R_i's and the G_i's are not sufficiently large, the differences $d_i = R_i - G_i$ may significantly contribute to define the value of ϑ. Then,

$$\begin{aligned} \vartheta &= 1/4[\sigma_3/\sigma_2^{3/2}]_N\{8R_1R_2R_3 + 4(d_1R_2R_3 + R_1d_2R_3 + R_1R_2d_3) \\ &\quad + R_1d_2d_3(2 - P_2 - P_3) + d_1R_2d_3(2 - P_1 - P_3) \\ &\quad + d_1d_2R_3(2 - P_2 - P_3) + d_1d_2d_3(1 + P_4)\} \\ &\approx 2[\sigma_3/\sigma_2^{3/2}]_N\left\{R_1R_2R_3 + 2R_3d_1d_2\left(1 + \frac{1}{\tan\delta_1\tan\delta_2}\right)\right. \\ &\quad + 2R_2d_1d_3\left(1 + \frac{1}{\tan\delta_1\tan\delta_3}\right) + 2R_1d_2d_3\left(1 + \frac{1}{\tan\delta_2\tan\delta_3}\right) \\ &\quad \left. + d_1d_2d_3\left(1 - \frac{1}{\tan\delta_1\tan\delta_2} + \frac{1}{\tan\delta_1\tan\delta_3} + \frac{1}{\tan\delta_2\tan\delta_3}\right)\right\}. \end{aligned} \quad (30)$$

As for γ, the factors $(\sin\delta_i)^{-1}$ take into account the larger signal provided by the anomalous scattering at higher $\sin\theta/\lambda$.

References

Dauter, Z., Wilson, K. S., Sieker, L. C., Meyer, J. & Moulis, J.-M. (1997). *Biochemistry*, **36**, 16065–16073.

Furey, W. Jr, Chandrasekhar, K., Dyda, F. & Sax, M. (1990). *Acta Cryst.* A**46**, 560–567.

Giacovazzo, C. (1983). *Acta Cryst.* A**39**, 585–592.

Giacovazzo, C., Cascarano, G. & Zheng, C.-D. (1988). *Acta Cryst.* A**44**, 45–51.

Giacovazzo, C., Ladisa, M. & Siliqi, D. (2002). *Acta Cryst.* A**58**, 598–604.

Giacovazzo, C., Siliqi, D. & De Caro, L. (2002). *Acta Cryst.* A**58**, 201–207.

Giacovazzo, C., Siliqi, D., Gonzalez-Platas, J., Hecht, H., Zanotti, G. & York, B. (1996). *Acta Cryst.* D**52**, 813–825.

Hauptman, H. (1982*a*). *Acta Cryst.* A**38**, 289–294.

Hauptman, H. (1982*b*). *Acta Cryst.* A**38**, 632–641.

Karle, J. (1984). *Acta Cryst.* A**40**, 4–11.

Sim, G. A. (1959). *Acta Cryst.* **12**, 813–816.

Sim, G. A. (1960). *Acta Cryst.* **13**, 511–512.

Walsh, M. A., Dementieva, I., Evans, O., Sanishili, R. & Joachimiak, A. (1999). *Acta Cryst.* D**55**, 1168–1173.